

WHAT IS CLAIMED IS:

1. A method for predicting biological functional activity and/or binding activity of an arbitrary protein, which comprises:

determining a total amino acid sequence frequency spectrum obtained by giving EIIP (Electron-ion interaction potential) index values to the amino acid residues of an arbitrary amino acid sequence originated in natural-type or non-natural-type and subjecting the resulting numerical value sequence (EIIP sequence) to discrete Fourier transformation (DFT), and

an active site frequency spectrum obtained by giving EIIP index values to the amino acids of an amino acid sequence region, which is composed of 2 to 64 amino acid residues present in the above arbitrary amino acid sequence and contains at least one known motif pertinent to an active site and subjecting the resulting EIIP sequence to DFT; and

selecting one or more characteristic frequency values derived from an active site of the protein from the cross-spectrum of the above total amino acid sequence frequency spectrum and the above active site frequency spectrum, and searching for one or more approximate frequency values of well-known characterized proteins similar to the characteristic frequency values described above.

2. The method for predicting biological functional activity and/or binding activity of an arbitrary protein

according to claim 1, wherein as the known motif as a signal of the active site, any one or more of GT, AS, GA, ID, TR, SR, LK, TXW, VXH, MXH, WXP, AXC, GXS (wherein G, T, A, S, I, D, R, L, K, W, V, H, M, P, C, and X mean glycine, threonine, alanine, serine, isoleucine, aspartic acid, arginine, leucine, lysine, tryptophan, valine, histidine, methionine, proline, cysteine, and any of 20 kinds of amino acids, respectively) and/or the reversed sequences thereof are employed.

3. A method for predicting biological functional activity and/or binding activity of an arbitrary protein, which comprises:

determining a total amino acid sequence frequency spectrum obtained by giving EIIP (Electron-ion interaction potential) index values to the amino acid residues of an arbitrary amino acid sequence originated in natural-type or non-natural-type and subjecting the resulting numerical value sequence (EIIP sequence) to discrete Fourier transformation (DFT), and

a total nucleotide sequence frequency spectrum obtained by giving EIIP index values to the nucleotide residues of a nucleotide sequence region academically corresponding to the above amino acid sequence and subjecting the resulting EIIP sequence to DFT; and

selecting one or more characteristic frequency values derived from the protein from the cross-spectrum of the above total amino acid sequence frequency spectrum and the above active site frequency spectrum, and

searching for one or more approximate frequency values of well-known characterized proteins similar to the characteristic frequency values described above.

4. A method for predicting biological functional activity and/or binding activity of an arbitrary nucleotide sequence, which comprises:

determining first total nucleotide sequence frequency spectrum obtained by giving EIIP (Electron-ion interaction potential) index values to the nucleotide residues of an arbitrary single-strand nucleotide sequence originated in natural-type or non-natural-type and subjecting the resulting numerical value sequence (EIIP sequence) to discrete Fourier transformation (DFT), and

second total nucleotide sequence frequency spectrum obtained by giving EIIP index values to the nucleotide residues of a nucleotide sequence which binds to the above nucleotide sequence through hydrogen bonding, and subjecting the resulting EIIP sequence to DFT; and

selecting one or more characteristic frequency values derived from the nucleotide sequence from the cross-spectrum of the above first total nucleotide sequence frequency spectrum and the above second nucleotide sequence frequency spectrum, and searching for one or more approximate frequency values of well-known characterized proteins similar to the characteristic frequency values described above.

5. A method for predicting biological functional activity and/or binding activity of an arbitrary amino acid sequence originated in natural-type or non-natural-type and other nucleotide sequence, which comprises:

determining at least two spectra of the following five spectra:

first spectrum of a total amino acid sequence frequency spectrum obtained by giving EIIP (Electron-ion interaction potential) index values to the amino acid residues of an arbitrary amino acid sequence originated in natural-type or non-natural-type and subjecting the resulting numerical value sequence (EIIP sequence) to discrete Fourier transformation (DFT),

second spectrum of an active site frequency spectrum obtained by giving EIIP index values to the amino acids of an amino acid sequence region, which is composed of 2 to 64 amino acid residues present in an arbitrary amino acid sequence and contains at least one known motif pertinent to an active site and subjecting the resulting EIIP sequence to DFT,

third spectrum of a total nucleotide sequence frequency spectrum obtained by giving EIIP index values to the nucleotide residues of a nucleotide sequence region academically corresponding to the amino acid sequence and subjecting the resulting EIIP sequence to DFT,

fourth spectrum of a total nucleotide sequence frequency spectrum obtained by giving EIIP index values

to the nucleotide residues of an arbitrary single-strand nucleotide sequence originated in natural-type or non-natural-type and subjecting the resulting EIIP sequence to DFT, and

fifth spectrum of a total nucleotide sequence frequency spectrum obtained by giving EIIP index values to the nucleotide residues of a complementary nucleotide sequence which binds to a nucleotide sequence through hydrogen bonding, and subjecting the resulting EIIP sequence to DFT; and

comparing with each spectrum.

6. The method for predicting biological functional activity and/or binding activity of an arbitrary amino acid sequence and other nucleotide sequence according to claim 5, wherein as the known motif as a signal of the active site, any one or more of GT, AS, GA, ID, TR, SR, LK, TXW, VXH, MXH, WXP, AXC, GXS (wherein G, T, A, S, I, D, R, L, K, W, V, H, M, P, C, and X mean glycine, threonine, alanine, serine, isoleucine, aspartic acid, arginine, leucine, lysine, tryptophan, valine, histidine, methionine, proline, cysteine, and any of 20 kinds of amino acids, respectively) and/or reverse sequences thereof are employed.

7. A method for predicting an active site of an arbitrary amino acid sequence or nucleotide sequence originated in natural-type or non-natural-type, which comprises:

determining at least two spectra of the following five spectra:

first spectrum of a total amino acid sequence frequency spectrum obtained by giving EIIP (Electron-ion interaction potential) index values to the amino acid residues of an arbitrary amino acid sequence originated in natural-type or non-natural-type and subjecting the resulting numerical value sequence (EIIP sequence) to discrete Fourier transformation (DFT),

second spectrum of an active site frequency spectrum obtained by giving EIIP index values to the amino acids of an amino acid sequence region, which is composed of 2 to 64 amino acid residues present in an arbitrary amino acid sequence and contains at least one known motif pertinent to an active site and subjecting the resulting EIIP sequence to DFT,

third spectrum of a total nucleotide sequence frequency spectrum obtained by giving EIIP index values to the nucleotide residues of a nucleotide sequence region academically corresponding to the amino acid sequence and subjecting the resulting EIIP sequence to DFT,

fourth spectrum of a total nucleotide sequence frequency spectrum obtained by giving EIIP index values to the nucleotide residues of an arbitrary single-strand nucleotide sequence originated in natural-type or non-natural-type and subjecting the resulting EIIP sequence to DFT, and

fifth spectrum of a total nucleotide sequence frequency spectrum obtained by giving EIIP index values to the nucleotide residues of a complementary nucleotide sequence which binds to a nucleotide sequence through hydrogen bonding, and subjecting the resulting EIIP sequence to DFT; and

comparing with each spectrum.

8. The method for predicting an active site of an arbitrary amino acid sequence or nucleotide sequence according to claim 7, wherein as the known motif as a signal of the active site, any one or more of GT, AS, GA, ID, TR SR, LK, TXW, VXH, MXH, WXP, AXC, GXS (wherein G, T, A, S, I, D, R, L, K, W, V, H, M, P, C, and X mean glycine, threonine, alanine, serine, isoleucine, aspartic acid, arginine, leucine, lysine, tryptophan, valine, histidine, methionine, proline, cysteine, and any of 20 kinds of amino acids, respectively) and/or reverse sequences thereof are employed.

9. A method for predicting biological functional activity and/or binding activity of an arbitrary amino acid sequence and/or an arbitrary nucleotide sequence, which comprises:

determining at least two spectra of the following five spectra:

first spectrum of a total amino acid sequence frequency spectrum obtained by giving EIIP (Electron-ion interaction potential) index values to the amino acid residues of an arbitrary amino acid sequence originated

in natural-type or non-natural-type and subjecting the resulting numerical value sequence (EIIP sequence) to discrete Fourier transformation (DFT),

second spectrum of an active site frequency spectrum obtained by giving EIIP index values to the amino acids of an amino acid sequence region, which is composed of 2 to 64 amino acid residues present in an arbitrary amino acid sequence and contains at least one known motif pertinent to an active site and subjecting the resulting EIIP sequence to DFT,

third spectrum of a total nucleotide sequence frequency spectrum obtained by giving EIIP index values to the nucleotide residues of a nucleotide sequence region academically corresponding to the amino acid sequence and subjecting the resulting EIIP sequence to DFT,

fourth spectrum of a total nucleotide sequence frequency spectrum obtained by giving EIIP index values to the nucleotide residues of an arbitrary single-strand nucleotide sequence originated in natural-type or non-natural-type and subjecting the resulting EIIP sequence to DFT, and

fifth spectrum of a total nucleotide sequence frequency spectrum obtained by giving EIIP index values to the nucleotide residues of a nucleotide sequence which binds to the nucleotide sequence through hydrogen bonding, and subjecting the resulting EIIP sequence to DFT; and comparing with each spectrum.

10. The method for predicting biological·functional activity and/or binding activity of an arbitrary amino acid sequence and/or an arbitrary nucleotide sequence according to claim 9, wherein as the known motif as a signal of the active site, any one or more of GT, AS, GA, ID, TR SR, LK, TXW, VXH, MXH, WXP, AXC, GXS (wherein G, T, A, S, I, D, R, L, K, W, V, H, M, P, C, and X mean glycine, threonine, alanine, serine, isoleucine, aspartic acid, arginine, leucine, lysine, tryptophan, valine, histidine, methionine, proline, cysteine, and any of 20 kinds of amino acids, respectively) and/or reverse sequences thereof are employed.

SUB A1
10026598 22701
11. A program wherein the method according to any one of claims 1 to 10 is constructed using a mathematical means, which realizes, on a computer, a function capable of predicting novel biological·functional and/or binding activity of desired protein, amino acid sequence, or nucleotide sequence.

12. The program according to claim 11, wherein the above mathematical means is Fourier analysis or wavelet analysis.

SUB A2
13. A storage medium readable on a computer, which stores a program wherein the method according to any one of claims 1 to 10 is constructed using a mathematical means, the program realizing, on a computer, a function capable of predicting novel biological·functional and/or binding activity of desired protein, amino acid sequence, or nucleotide sequence.

14. A binding mode of at least two kinds of arbitrary proteins (or amino acid sequences, nucleotide sequences), which is predicted by the method according to any one of claims 1 to 10.

SUB A2
cont

15. An application of biological functional activity of an arbitrary protein or nucleotide sequence predicted by the method according to any one of claims 1 to 10,

the activity being employed for at least one selected from a pesticide as a therapeutic agent, a medicament as a therapeutic agent, prevention of a hereditary disease, diagnosis of a hereditary disease, prevention of a pestilence, and diagnosis of a pestilence.

10026698-122701